

Tail dependence of recursive max-linear models with regularly varying noise variables

Nadine Gissibl*

Claudia Klüppelberg*

Moritz Otto*

January 26, 2017

Abstract

We investigate multivariate regularly varying random vectors with discrete spectral measure induced by a directed acyclic graph (DAG). The tail dependence coefficient measures extreme dependence between two vector components, and we investigate how the matrix of tail dependence coefficients can be used to identify the full dependence structure of the random vector on a DAG or even the DAG itself. Furthermore, we estimate the distributional model by the matrix of empirical tail dependence coefficients. Here we assume that we do not know the DAG, but observe only the multivariate data. From these observations we want to infer the causal dependence structure in the data.

AMS 2010 Subject Classifications: primary: 62G32, 60G70, 05C75 secondary: 62-09, 65S05

Keywords: DAG; directed acyclic graph; exponent measure; Markov graph; max-linear distribution; max-stable model; regular variation; structural equation model; extreme value theory; tail dependence coefficient.

1 Max-linear models on directed acyclic graphs

Extreme value (or max-stable) distributions occur naturally as limit models for centered and scaled maxima. Multivariate max-stable models and their domains of attraction have been investigated from a probabilistic and statistical point of view; see e.g. [2, 4, 15, 16].

We investigate multivariate regularly varying random vectors on graphs, which are in the maximum domain of attraction of a Fréchet distribution. The dependence between the components is modelled by the exponent measure, which in our case is induced by a graph. More precisely, the dependence structure of a random vector \mathbf{X} is given by a DAG $\mathcal{D} = (V, E)$ with node set $V = \{1, \dots, d\}$ and edge set $E = \{(k, i) \text{ for all } k \in \text{pa}(i)\}$, where $\text{pa}(i)$ denotes the *parents of node i*. As is commonly done, we identify the nodes with the components of the vector \mathbf{X} . The necessary background on graphical models can be found in Edwards [7] and Lauritzen [12] (also Lauritzen [11] provides a very readable summary), for structural equation models we refer to Bollen [3] and Pearl [13].

Throughout we assume as data generating mechanism a *recursive max-linear structural equation model*, which has representation

$$X_i = \bigvee_{k \in \text{pa}(i)} c_k^i X_k \vee c_i^i Z_i, \quad i = 1, \dots, d, \quad (1.1)$$

*Technische Universität München, Center for Mathematical Sciences, 85748 Garching, Boltzmannstrasse 3, Germany, e-mail: n.gissibl@tum.de, cklu@tum.de, moritz.otto@kit.edu

where Z_1, \dots, Z_d are iid random variables with absolutely continuous distribution and support $\mathbb{R}_+ = (0, \infty)$, and the $c_k^i > 0$ for all $k \in \text{Pa}(i) := \text{pa}(i) \cup \{i\}$. Extending classical notation slightly, c_k^i are called the *edge weights* of \mathcal{D} . This model is max-linear (cf. Theorem 2.2 of [8]) with representation

$$X_i = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j, \quad i = 1, \dots, d, \quad (1.2)$$

where $\text{An}(i) := \text{an}(i) \cup \{i\}$ and $\text{an}(i)$ are the *ancestors of node i* . We call the vector \mathbf{X} a *recursive max-linear (ML) model on \mathcal{D}* .

The coefficients b_{ji} are given as maxima of products of the edge weights c_j^i along paths from j to i . More precisely, for a path $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$ of length n from j to i , we define

$$d_{ji}(p) := c_{k_0}^{k_1} \prod_{l=0}^{n-1} c_{k_l}^{k_{l+1}} \quad \text{and} \quad b_{ji} := \bigvee_{p \in P_{ji}} d_{ji}(p), \quad (1.3)$$

where P_{ji} denotes all paths from j to i ; furthermore, we define $b_{ii} := c_i^i$ and $b_{ji} := 0$ for all $i \in V$ and $j \in V \setminus \text{An}(i)$. For $i \in V$ and $j \in \text{an}(i)$ we call a path p from j to i *max-weighted*, if $b_{ji} = d_{ji}(p)$. We summarize these coefficients in the *max-linear (ML) coefficient matrix*

$$B = (b_{ij})_{d \times d}. \quad (1.4)$$

Throughout this paper we assume that the distributions of the noise variables are in the maximum domain of attraction of a standard α -Fréchet distribution given for $\alpha > 0$ by

$$\Phi_\alpha(x) = \exp\{-x^{-\alpha}\} \mathbf{1}_{(0, \infty)}(x).$$

Thus, a generic noise variable Z is regularly varying with index $-\alpha$ for some $\alpha > 0$. However, the dependence structure, which is introduced by the DAG, will be max-linear not only in the max-stable limit, but already in the original regularly varying vector (see Remark 2.6 for obvious extensions). By max-linearity, we can write down the full dependence structure of the model explicitly.

In this paper we investigate the *tail dependence coefficient* defined for equally distributed random variables X_i, X_j as

$$\chi(i, j) = \chi(j, i) = \lim_{u \rightarrow \infty} P(F_i(X_i) > u \mid F_j(X_j) > u).$$

We shall show that in a recursive ML model on a DAG with regularly varying noise variables the tail dependence coefficient is positive, if and only if the two nodes i and j have joint ancestors.

The goal of this paper is two-fold. Firstly, we investigate how far the matrix χ of all tail dependence coefficients can identify the dependence structure of the recursive ML model \mathbf{X} on a DAG or even the DAG \mathcal{D} itself. Secondly, we estimate χ with the goal of statistical estimation of such graphical dependence structures. We assume that we do not know the DAG, but observe only data of the vector \mathbf{X} . From these observations we want to infer the DAG and the ML coefficient matrix B and the edge weight matrix C ; i.e., the causal ML structure of the data.

More precisely, in section 2 we provide some preliminary results and definitions, both in the context of extreme value theory and graphical models. In particular, we derive the exponent measure, which characterizes the dependence in regularly varying models, which is here discrete,

since induced by the DAG. Furthermore, we introduce the recursive ML max-weighted model, which will be our leading model in what follows. The tail dependence coefficient is introduced in section 3 and the link between the dependence structure of \mathbf{X} and the positivity of χ is discussed. The important question of identifiability of the recursive ML model \mathbf{X} from its tail dependence coefficients is investigated in section 4. Moreover, we explain the prominent role of the initial nodes V_0 of the DAG \mathcal{D} for the identifiability of the model. In section 5, the previous theory is applied to derive an identifiability result for the important class of recursive ML max-weighted models. Section 6 is devoted to the statistical estimation of the tail dependence coefficient matrix. We consider an empirical estimator of χ and give precise conditions for asymptotic multivariate normality. The result holds for more general models, but for a recursive ML model we give the asymptotic covariance matrix explicitly in terms of the max-linear coefficient matrix (1.4).

The following notation will be used throughout the paper.

For a node $i \in V$ the sets $\text{an}(i)$, $\text{pa}(i)$, and $\text{de}(i)$ denote the *ancestors*, *parents*, and *descendants* of i with respect to \mathcal{D} . Furthermore, we use the notation $\text{An}(i) = \text{an}(i) \cup \{i\}$, $\text{Pa}(i) = \text{pa}(i) \cup \{i\}$, and $\text{De}(i) = \text{de}(i) \cup \{i\}$. We denote by $j \rightarrow i$ an edge and by $[j \Rightarrow i]$ any path from i to j .

For any DAG \mathcal{D} we call the *initial nodes* $V_0 := \{i \in V : \text{an}(i) = \emptyset\}$ and the *terminal nodes* $V_\infty := \{i \in V : \text{de}(i) = \emptyset\}$.

In general, for arbitrary (possibly random) $a_i \geq 0$ we set $\bigvee_{i \in \emptyset} a_i = 0$, $\bigwedge_{i \in \emptyset} a_i = \infty$, $\sum_{i \in \emptyset} a_i = 0$, and $\prod_{i \in \emptyset} a_i = 1$.

2 Preliminaries

2.1 Distributional properties of \mathbf{X}

Throughout this paper we assume the iid noise variables Z_1, \dots, Z_n to be regularly varying. More precisely, a generic noise variable Z has distribution tail $\overline{F}_Z(x) = 1 - F_Z(x) = L(x)x^{-\alpha}$ for $x > 0$ for a slowly varying function L and index of regular variation $\alpha > 0$. We abbreviate this as $Z \in \mathcal{R}(-\alpha)$ or $F_Z \in \mathcal{R}(-\alpha)$. For details on multivariate regular variation and extreme value theory we refer to Resnick [15, 16].

Since $\mathbf{Z} := (Z_1, \dots, Z_d)$ has independent regularly varying components, its exponent measure is concentrated on the axes. Like every exponent measure it is uniquely defined by its definition on complements of the sets $[\mathbf{0}, \mathbf{x}]$, denoted by $[\mathbf{0}, \mathbf{x}]^c$ for $\mathbf{x} = (x_1, \dots, x_d) > (0, \dots, 0) = \mathbf{0}$. More precisely, for a_n chosen such that $n\mathbb{P}(Z > a_n) \xrightarrow{n \rightarrow \infty} 1$, we find (taking order relations componentwise)

$$\mu_Z([\mathbf{0}, \mathbf{x}]^c) := n\mathbb{P}(a_n^{-1}\mathbf{Z} \in [\mathbf{0}, \mathbf{x}]^c) = n(1 - \mathbb{P}(\mathbf{Z} \leq a_n\mathbf{x})) \xrightarrow{n \rightarrow \infty} -\log \prod_{k=1}^d \Phi_\alpha(x_k) = \sum_{k=1}^d x_k^{-\alpha}. \quad (2.1)$$

Proposition 2.1. *Let \mathbf{X} be a recursive ML model on a DAG. Then the distribution function of $\mathbf{X} = (X_1, \dots, X_d)$ is in the maximum domain of attraction of*

$$G(\mathbf{x}) = \exp \left\{ - \sum_{k=1}^d \bigvee_{i \in \text{De}(k)} \left(\frac{b_{ki}}{x_i} \right)^\alpha \right\} \mathbf{1}_{(0, \infty)}(\mathbf{x}). \quad (2.2)$$

Hence, \mathbf{X} has exponent measure

$$\mu_X([\mathbf{0}, \mathbf{x}]^c) = \sum_{k=1}^d \bigvee_{i \in \text{De}(k)} \left(\frac{b_{ki}}{x_i} \right)^\alpha. \quad (2.3)$$

In particular, the one-dimensional and bivariate marginal distribution functions are given for $i, j = 1, \dots, d$, by

$$G_i(x) = \exp \left\{ -x^{-\alpha} \sum_{k \in \text{An}(i)} b_{ki}^\alpha \right\} \mathbf{1}_{(0, \infty)}(x),$$

$$G_{ij}(x_i, x_j) = \exp \left\{ - \sum_{k \in \text{An}(i) \cap \text{An}(j)} \left(\frac{b_{ki}}{x_i} \right)^\alpha \wedge \left(\frac{b_{kj}}{x_j} \right)^\alpha \right\} \mathbf{1}_{(0, \infty) \times (0, \infty)}(x_i, x_j),$$

Proof. With $\mathbf{X} = (X_1, \dots, X_d)$ and $\mathbf{x} = (x_1, \dots, x_d)$ as above, we have by (1.2),

$$\begin{aligned} n(1 - \mathbb{P}(\mathbf{X} \leq a_n \mathbf{x})) &= n \left(1 - \mathbb{P} \left(\bigvee_{k=1}^d b_{ki} Z_k \leq a_n x_i, i = 1, \dots, d \right) \right) \\ &= n \left(1 - \mathbb{P} \left(Z_k \leq a_n \bigwedge_{i \in \text{De}(k)} \frac{x_i}{b_{ki}}, k = 1, \dots, d \right) \right). \end{aligned}$$

By (2.1), we conclude that $n(1 - \mathbb{P}(\mathbf{X} \leq a_n \mathbf{x})) \xrightarrow{n \rightarrow \infty} \sum_{k=1}^d \bigvee_{i \in \text{De}(k)} \left(\frac{b_{ki}}{x_i} \right)^\alpha$. This implies (2.2). The marginal distribution functions are obtained by letting all other arguments of G tend to ∞ . \square

Remark 2.2. The one-dimensional marginal distributions and also the exponent measure depend on the max-linear coefficients. This can cause difficulties, when the marginal distributions differ; i.e., when $\sum_{k \in \text{An}(i)} b_{ki} \neq \sum_{k \in \text{An}(j)} b_{kj}$ for $i, j \in V$. Thus we follow standard practice and calculate the exponent measure after standardization of the marginals to standard unit Fréchet (cf. [15], Section 5.4.1.). This requires the definition of the *generalized inverse* of an increasing function f given by $f^\leftarrow(y) := \inf\{x : f(x) \geq y\}$. If f is strictly increasing, then f^\leftarrow coincides with the analytic inverse. Then a standardized version of G is obtained as follows:

$$G_*(\mathbf{x}) := G \left(\left(\frac{1}{-\log G_1} \right)^\leftarrow(x_1), \dots, \left(\frac{1}{-\log G_d} \right)^\leftarrow(x_d) \right) \mathbf{1}_{(0, \infty)}(\mathbf{x}) \quad (2.4)$$

with exponent measure given for $\mathbf{x} = (x_1, \dots, x_d) > (0, \dots, 0)$ by

$$\begin{aligned} \mu_*([\mathbf{0}, \mathbf{x}]^c) &:= -\log G_*(\mathbf{x}) = -\log G \left((x_1 \sum_{l=1}^d b_{l1}^\alpha)^{1/\alpha}, \dots, (x_d \sum_{l=1}^d b_{ld}^\alpha)^{1/\alpha} \right) \\ &= \sum_{k=1}^d \bigvee_{i \in \text{De}(k)} \frac{b_{ki}^\alpha}{\sum_{l=1}^d b_{li}^\alpha} x_i^{-1}. \end{aligned} \quad (2.5)$$

\square

Remark 2.3. The dependence structure, which is introduced by the DAG, is max-linear not only in the max-stable limit, but already in the original regularly varying vector \mathbf{X} . This rather limited dependence structure can be generalized naturally within the framework of regular variation.

The extent of the possible generalization can be best understood, when considering an equivalent representation of the dependence in a regularly varying vector. According to Resnick [16], Theorem 6.1, $\mathbf{X} \in \mathbb{R}^d$ is multivariate regularly varying, if and only if there exists a random vector $\Theta \in \mathcal{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$ such that for $t > 0$:

$$\frac{\mathbb{P}(\|\mathbf{X}\| > tx, \mathbf{X}/\|\mathbf{X}\| \in \cdot)}{\mathbb{P}(\|\mathbf{X}\| > x)} \xrightarrow{w} t^{-\alpha} \mathbb{P}(\Theta \in \cdot), \quad x \rightarrow \infty, \quad (2.6)$$

where \xrightarrow{w} denotes weak convergence of measures and $\|\cdot\|$ is any norm in \mathbb{R}^d . The probability distribution S of Θ is called the *spectral measure* of \mathbf{X} . From (2.6) we find immediately that the

dependence structure of \mathbf{X} is for moderate values of $\|\mathbf{X}\|$ arbitrary, only when the norm of the vector becomes large, the dependence structure becomes that of Θ . By Proposition 6.4 of [16] the exponent measure and spectral measure are related by a polar coordinate transform giving

$$\mu([\mathbf{0}, \mathbf{x}]^c) = c \int_{\mathcal{S}_+^{d-1}} \bigvee_{i=1}^d \left(\frac{a_i}{x_i} \right)^\alpha S(d\mathbf{a}), \quad \mathbf{x} > \mathbf{0},$$

for some constant $c > 0$ and $\mathbf{a} = (a_1, \dots, a_d)$. If S had been a continuous measure, $\mathbf{a} = (a_1, \dots, a_d)$ could take any value on the positive unit sphere \mathcal{S}_+^{d-1} in \mathbb{R}^d . In our setting, however, the dependence structure of \mathbf{X} becomes discrete in the limit, such that each component a_i can only take values b_{ki} for $k = 1, \dots, d$.

When statistical estimation is based on the extreme values of a sample, as in this paper when estimating the tail dependence coefficients, the restriction to the (limiting) discrete dependence provides a sufficient model. Furthermore, the slightly simpler model induced by the DAG for all values of the data allows for more concise notation and makes the new ideas given by the DAG dependence structure more transparent. \square

2.2 Structural properties

The following lemma describes a property of the max-linear coefficient matrix B . Note that by Proposition 2.1 the condition on the ML coefficients implies that the one-dimensional marginal distributions are standard unit Fréchet.

Lemma 2.4. *Let $B = (b_{ij})_{d \times d}$ be the ML coefficient matrix of a recursive ML model on a DAG such that $\sum_{l \in \text{An}(i)} b_{li}^\alpha = 1$ for all $i \in V$. Then for all $i, j \in V$,*

$$b_{ji} < b_{jj}.$$

Proof. Note that $b_{jj} > 0$ for all $j \in V$. If $b_{ji} = 0$, which is by (1.3) the case if $j \in V \setminus \text{An}(i)$, then it obviously holds that $b_{ji} < b_{jj}$. Assume now that $j \in \text{An}(i)$. Using $b_{li} \geq \frac{b_{lj}b_{ji}}{b_{jj}}$ for all $l \in \text{An}(j)$ (cf. Lemma 2.10 of [8]) and the fact that $\sum_{l \in \text{An}(j)} b_{lj} = 1$, we obtain

$$1 = \sum_{l \in \text{An}(i)} b_{li} = \sum_{l \in \text{An}(j)} b_{li} + \sum_{l \in \text{An}(i) \setminus \text{An}(j)} b_{li} \geq \frac{b_{ji}}{b_{jj}} + \sum_{l \in \text{An}(i) \setminus \text{An}(j)} b_{li},$$

Since $b_{li} > 0$ for all $l \in \text{An}(i) \setminus \text{An}(j)$, we have that $\frac{b_{ji}}{b_{jj}} < 1$, equivalently $b_{ji} < b_{jj}$. \square

From (1.3) we know that not all paths may be relevant for obtaining the maximum in definition (1.1) of the model. In the next definition we summarize the relevant paths, where the maximum is realised.

Definition 2.5. [Gissibl and Klüppelberg [8], Definition 2.6]

For $i \in V$ and $j \in \text{an}(i)$ we call a path $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$ from j to i *max-weighted*, if p realises the maximum in (1.3); i.e., if $b_{ji} = d_{ji}(p) = c_{k_0}^{k_0} \prod_{l=0}^{k_1-1} c_{k_l}^{k_{l+1}}$. \square

The DAG with the minimal number of edges corresponding to a given reachability matrix $R = \text{sgn}(B)$ is defined as follows.

Definition 2.6. A DAG $\mathcal{D}^{\text{tr}} = (V, E^{\text{tr}})$ is called *transitive reduction* of \mathcal{D} , if

- (i) for all $i, j \in V$, \mathcal{D}^{tr} has a path from j to i if and only if \mathcal{D} has a path from j to i , and
- (ii) there is no DAG with less edges satisfying condition (a). \square

As shown in Aho et al. [1], the transitive reduction of a finite DAG \mathcal{D} is unique and also a subgraph of \mathcal{D} .

We define an analog of the transitive reduction of a DAG \mathcal{D} in the context of recursive ML models. Intuitively, we cannot dispose of those paths, which are max-weighted in the sense of Definition 2.5. Thus, for a given ML coefficient matrix B the DAG \mathcal{D}^B with minimal number of edges, such that \mathbf{X} is a recursive ML model on a DAG corresponding to B , has in general more edges than the transitive reduction DAG.

Definition 2.7. A DAG \mathcal{D}^B is called *max-linear reduction* if for all $i, j \in V$, \mathcal{D}^B has an edge $j \rightarrow i$ if and only if this is the only max-weighted path from j to i in \mathcal{D} . \square

It has been shown in Theorem 3.5 of Gissibl and Klüppelberg [8] that \mathcal{D}^B is unique and also a subgraph of \mathcal{D} . Moreover, the recursive ML model \mathbf{X} with ML coefficient matrix B is minimal with respect to \mathcal{D}^B , which can be characterized as

$$\mathcal{D}^B = (V, E^B) := (V, \{(k, i) \in E : b_{ki} > \bigvee_{l \in \text{de}(k) \cap \text{pa}(i)} \frac{b_{kl}b_{li}}{b_{ll}}\}). \quad (2.7)$$

With the next definition we introduce our leading example of a recursive ML model that will be further investigated in Section 5.

Definition 2.8. Let \mathbf{X} be a recursive ML model on a DAG with ML coefficient matrix $B = (b_{ij})_{d \times d}$. Assume that all paths are max-weighted; i.e., for arbitrary nodes $i, j \in V$ and all paths $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$, independent of the specific path p ,

$$b_{ji} = c_{k_0}^{k_0} c_{k_0}^{k_1} \dots c_{k_{n-2}}^{k_{n-1}} c_{k_{n-1}}^{k_n} = d_{ji}(p).$$

Then we call \mathbf{X} *recursive ML max-weighted model*. Since every path in \mathcal{D} is max-weighted, we observe from (1.3) by comparing two existing paths $[j \Rightarrow i]$ and $[j \Rightarrow k_m \Rightarrow i]$ that

$$b_{ji} = \frac{b_{jk_m} b_{k_m i}}{b_{k_m k_m}} \quad \text{for } m = 1, \dots, n-1. \quad (2.8)$$

\square

Example 2.9. [Polytree]

Let \mathbf{X} be a recursive ML model relative to a polytree \mathcal{D} ; i.e., \mathcal{D} has no cycles, equivalently, the underlying undirected graph is a tree (cf. Koller and Friedman [10], Definition 2.2). Since there exists at most one path between every pair of nodes, all paths must be max-weighted. \square

The following example shows that for every DAG we can find a recursive ML max-weighted model.

Example 2.10. [Homogeneous model]

Let \mathbf{X} be a recursive ML structural equation model as in (1.1) defined by

$$X_i := \frac{1}{|\text{An}(i)|^{1/\alpha}} \left(\bigvee_{k \in \text{pa}(i)} |\text{An}(k)|^{1/\alpha} X_k \vee Z_i \right), \quad i = 1, \dots, d.$$

Let $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$ be a path of length n from j to i . Then the coefficient $d_{ji}(p)$ from (1.3) is given by $d_{ji}(p) = |\text{An}(i)|^{-1/\alpha}$ such that

$$X_i = \frac{1}{|\text{An}(i)|^{1/\alpha}} \bigvee_{j \in \text{An}(i)} Z_j, \quad i = 1, \dots, d;$$

i.e., the ML coefficient matrix is $B = (b_{ji})_{d \times d} = \left(\frac{1}{|\text{An}(i)|^{1/\alpha}} \mathbb{1}_{\text{An}(i)}(j) \right)_{d \times d}$. \square

Lemma 2.11. *Let \mathbf{X} be a recursive ML max-weighted model with respect to a DAG \mathcal{D} . Then $\mathcal{D}^B = \mathcal{D}^{\text{tr}}$.*

Proof. Since \mathcal{D}^{tr} is a subgraph of \mathcal{D}^B and minimal, we only have to consider $i \in V$ and $k \in \text{pa}(i) \setminus \text{pa}^{\text{tr}}(i)$, where $\text{pa}^{\text{tr}}(i)$ denote the transitive parents of i . Since every path from k to i is max-weighted, we have also a max-weighted path which contains a node of $\text{pa}(i) \cap \text{de}(k)$. Thus we know from Lemma 2.10(a) of [8] that

$$b_{ki} = \bigvee_{l \in \text{pa}(i) \cap \text{de}(k)} \frac{b_{kl}b_{li}}{b_{ll}}.$$

This contradicts (2.7). \square

3 Extreme dependence measures

In what follows we assume that $\mathbf{X} = (X_1, \dots, X_d)$ is a recursive ML structural equation model relative to a DAG given by (1.2) with iid noise variables $Z_k \in \mathcal{R}(-\alpha)$ for $k = 1, \dots, d$. Denote by F the joint distribution function of \mathbf{X} and by F_i the marginal distribution function of X_i for $i = 1, \dots, d$.

Various functions and coefficients have been suggested to describe the extreme dependence of a random vector in a different and possibly simpler way than by the full exponent measure (2.3) or its standardised version (2.5). We focus on the following.

Definition 3.1. For nodes $i, j \in V$ define the *tail dependence coefficient* between X_i and X_j by

$$\chi(i, j) := \lim_{u \rightarrow 1} \mathbb{P}(F_i(X_i) > u \mid F_j(X_j) > u). \quad (3.1)$$

We summarize all $\chi(i, j)$ in a matrix $\chi := (\chi(i, j))_{d \times d}$ and call χ *tail dependence coefficient matrix*. \square

Such measures have been defined and used in the context of multivariate distributions (see for example Beirlant et al. [2, Section 9.5.1]). Moreover, they are usually restricted to distributions with equal marginals, or applied after transforming different marginals as we also suggest in Remark 2.2.

In the following theorem we express the tail dependence coefficient and its multivariate extensions explicitly in terms of the ML coefficient matrix B .

Theorem 3.2. *For $i, j \in V$ define*

$$\bar{b}_{ij} := \frac{b_{ij}^\alpha}{\sum_{l=1}^d b_{lj}^\alpha}.$$

Let $i_1, \dots, i_l \in V$ be an arbitrary index set. Then

$$\begin{aligned} & \lim_{u \rightarrow 1} \mathbb{P}(F_{i_1}(X_{i_1}) > u, \dots, F_{i_l}(X_{i_l}) > u \mid F_{i_1}(X_{i_1}) > u) \\ &= \lim_{u \rightarrow 1} \frac{1}{1-u} \mathbb{P}(F_{i_1}(X_{i_1}) > u, \dots, F_{i_l}(X_{i_l}) > u) = \sum_{k \in \text{An}(i_1) \cap \dots \cap \text{An}(i_l)} \bigwedge_{m=1}^l \bar{b}_{ki_m}. \end{aligned} \quad (3.2)$$

In particular, the tail dependence coefficient $\chi(i, j)$ between X_i and X_j is given by

$$\chi(i, j) = \sum_{k \in \text{An}(i) \cap \text{An}(j)} \bar{b}_{ki} \wedge \bar{b}_{kj}. \quad (3.3)$$

Proof. For $i, j \in V$, eq. (3.2) is equivalent to (3.3). The general formula (3.2) can be derived from the multivariate distribution function (2.2) by the usual inclusion-exclusion argument. For notational ease we only present a proof of (3.3). By [2], Section 9.5.1, for $i, j \in V$ (3.2) is equal to the tail dependence coefficient $\chi(i, j)$, which is given by

$$\chi(i, j) = 2 - \lim_{u \rightarrow 1} \frac{1 - F(\infty, \dots, \infty, F_i^{\leftarrow}(u), \infty, \dots, \infty, F_j^{\leftarrow}(u), \infty, \dots, \infty)}{1 - u}.$$

Substituting $t := \frac{1}{1-u}$ and observing that for the generalized inverse $F_i^{\leftarrow}(1 - \frac{1}{t}) = \left(\frac{1}{1-F_i}\right)^{\leftarrow}(t)$ holds, we have with $\mathbf{t} = (t_1, \dots, t_d)$ for $F_*(\mathbf{t}) := F\left(\left(\frac{1}{1-F_1}\right)^{\leftarrow}(t_1), \dots, \left(\frac{1}{1-F_d}\right)^{\leftarrow}(t_d)\right)$ that

$$\chi(i, j) = 2 - \lim_{t \rightarrow \infty} t(1 - F_*(t(e_i \wedge e_j))),$$

where the vector e_i is 1 at position i and ∞ elsewhere, and the minimum is understood componentwise. By [15], Prop. 5.15(b), F_* is in the maximum domain of attraction of the distribution function G_* as given in Remark 2.2. Using Prop. 5.15(a) and Eq. (5.38) of [15], we conclude that

$$t(1 - F_*(t(e_i \wedge e_j))) \xrightarrow{t \rightarrow \infty} -\log G_*(e_i \wedge e_j) = \mu_*([0, e_i \wedge e_j]^c).$$

This yields

$$\chi(i, j) = 2 - \sum_{k=1}^d \frac{b_{ki}^\alpha}{\sum_{l=1}^d b_{li}^\alpha} \vee \frac{b_{kj}^\alpha}{\sum_{l=1}^d b_{lj}^\alpha}.$$

Since

$$\sum_{k=1}^d \frac{b_{ki}^\alpha}{\sum_{l=1}^d b_{li}^\alpha} \vee \frac{b_{kj}^\alpha}{\sum_{l=1}^d b_{lj}^\alpha} + \sum_{k=1}^d \frac{b_{ki}^\alpha}{\sum_{l=1}^d b_{li}^\alpha} \wedge \frac{b_{kj}^\alpha}{\sum_{l=1}^d b_{lj}^\alpha} = 2,$$

and $b_{ki} \wedge b_{kj} = 0$ for $k \in V \setminus (\text{An}(i) \cap \text{An}(j))$ (cf. (1.3)), we finally obtain

$$\chi(i, j) = \sum_{k \in \text{An}(i) \cap \text{An}(j)} \frac{b_{ki}^\alpha}{\sum_{l=1}^d b_{li}^\alpha} \wedge \frac{b_{kj}^\alpha}{\sum_{l=1}^d b_{lj}^\alpha}.$$

□

Remark 3.3. By Theorem 3.3 of [8], the matrix $\overline{B} = (\overline{b}_{ij})_{d \times d}$ is again a ML coefficient matrix of a recursive ML model. □

Corollary 3.4. *The following are equivalent.*

- (a) $\chi(i, j) = 0$
- (b) $\text{An}(i) \cap \text{An}(j) = \emptyset$
- (c) X_i and X_j are independent

Proof. (a) \iff (b) is immediate by (3.3).

(b) \iff (c) holds by definition of X_i and X_j . Since $X_i := \bigvee_{k \in \text{An}(i)} b_{ki} Z_k$ with independent noise variables Z_k , X_i and X_j are independent if and only if $\text{An}(i) \cap \text{An}(j) = \emptyset$. □

Remark 3.5. In the more general framework of Remark 2.6, parts (b) and (c) of Cor. 3.4 would have to be replaced by:

- (b') The limiting distribution can be presented on a DAG such that $\text{An}(i) \cap \text{An}(j) = \emptyset$
(c') X_i and X_j are asymptotically independent. \square

The following result states some useful properties of the tail dependence coefficients.

Proposition 3.6. *Let $i \in V$.*

- (a) *For $j \in \text{an}(i)$ we have $0 < \chi(i, j) \leq \sum_{k \in \text{An}(j)} \bar{b}_{ki} < 1$.*
(b) *For $j \in \text{An}(i)$ we have $\chi(i, j) \geq \frac{\bar{b}_{ji}}{b_{jj}}$ with equality if and only if all paths from $k \in \text{An}(j)$ to i passing through j are max-weighted.*

Let V_0 be the set of initial nodes of \mathcal{D} , then

- (c) *For $j \in \text{An}(i) \cap V_0$ we have $\chi(i, j) = \bar{b}_{ji}$.*
(d) *For $j \in \text{An}(i) \cap V_0$ we have*

$$\bigvee_{k \in \text{De}(j) \cap \text{An}(i)} \chi(j, k) \chi(k, i) \geq \chi(j, i),$$

with equality if for every $k \in \text{De}(j) \cap \text{An}(i)$, every path from $l \in \text{An}(k)$ to i passing through k is max-weighted.

Proof. (a) We obtain from (3.3), since $\bar{b}_{ki} \wedge \bar{b}_{kj} = 0$ for $k \notin \text{An}(j)$,

$$\chi(i, j) = \sum_{k=1}^d \bar{b}_{ki} \wedge \bar{b}_{kj} = \sum_{k \in \text{An}(j)} \bar{b}_{ki} \wedge \bar{b}_{kj} \leq \sum_{k \in \text{An}(j)} \bar{b}_{ki} < \sum_{k \in \text{An}(i)} \bar{b}_{ki} = 1,$$

where the inequality is due to $\text{An}(j) \not\subseteq \text{An}(i)$.

(b) From Lemma 2.10 of [8] we know that $b_{ki} \geq \frac{b_{kj} b_{ji}}{b_{jj}}$ for all $k \in \text{An}(j)$. Remark 3.3 and Lemma 2.4 imply $\frac{\bar{b}_{ji}}{b_{jj}} < 1$. Hence,

$$\chi(i, j) = \sum_{k \in \text{An}(j)} \bar{b}_{ki} \wedge \bar{b}_{kj} \geq \sum_{k \in \text{An}(j)} \frac{\bar{b}_{kj} \bar{b}_{ji}}{\bar{b}_{jj}} \wedge \bar{b}_{kj} = \frac{\bar{b}_{ji}}{\bar{b}_{jj}} \sum_{k \in \text{An}(j)} \bar{b}_{kj} = \frac{\bar{b}_{ji}}{\bar{b}_{jj}}.$$

The second statement follows from the fact that a path from $k \in \text{An}(j)$ to i through j is max-weighted if and only if $b_{ki} = \frac{b_{kj} b_{ji}}{b_{jj}}$.

- (c) Since $\sum_{l \in \text{An}(j)} \bar{b}_{lj} = 1$ implies $\bar{b}_{jj} = 1$ for all $j \in V_0$, the statement is a special case of (b).
(d) Since $j \in \text{De}(j) \cap \text{An}(i)$ and $\chi(j, j) = 1$, it holds that

$$\bigvee_{k \in \text{De}(j) \cap \text{An}(i)} \chi(j, k) \chi(k, i) \geq \chi(j, j) \chi(j, i) = \chi(j, i).$$

Let $k \in \text{De}(j) \cap \text{An}(i)$. If all paths from $l \in \text{An}(k)$ to i through k are max-weighted, we have by parts (b) and (c) that $\chi(j, k) = \bar{b}_{jk}$ and $\chi(k, i) = \frac{\bar{b}_{ki}}{b_{kk}}$, implying

$$\bigvee_{k \in \text{De}(j) \cap \text{An}(i)} \chi(j, k) \chi(k, i) = \bigvee_{k \in \text{De}(j) \cap \text{An}(i)} \bar{b}_{jk} \frac{\bar{b}_{ki}}{b_{kk}} = \bar{b}_{ji} = \chi(j, i),$$

where the equality $\bigvee_{k \in \text{De}(j) \cap \text{An}(i)} \bar{b}_{jk} \frac{b_{ki}}{b_{kk}} = b_{ji}$ is due to Lemma 2.10 of [8]. \square

Example 3.7. Consider the following DAGs \mathcal{D}_1 (left) and \mathcal{D}_2 (right).



(1) In the second statement of Lemma 3.6(d), the assumption that all paths from $l \in \text{An}(k)$ to i through k are max-weighted cannot be dropped: consider a max-linear model with $\alpha = 1$ and $\sum_{k=1}^d b_{ki} = 1$ for all $i \in V$ on the left hand DAG \mathcal{D}_1 with $c_1^3 < c_1^4 = b_{14}$. Since $\overline{B} = B$ and $c_3^4 < 1$, we have by Lemma 3.6:

$$\chi(2, 3) = c_2^3, \quad \chi(3, 4) = c_3^4, \quad \text{and} \quad \chi(2, 4) = c_2^3 c_3^4.$$

Hence, $\chi(2, 3)\chi(3, 4) > c_2^3 c_3^4 = \chi(2, 4)$. Since the path $[1 \rightarrow 3 \rightarrow 4]$ is not max-weighted, equality as in Lemma 3.6(d) does not hold.

(2) $\chi(j, k)\chi(k, i) = \chi(j, i)$ does not imply that the path $[j \rightarrow k \rightarrow i]$ is max-weighted: consider the max-linear model with $\alpha = 1$ on the right hand DAG \mathcal{D}_2 with weight and ML coefficient matrices

$$C = \begin{pmatrix} 1 & 0 & 0.1 & 0.085 \\ 0 & 1 & 0.8 & 0.5 \\ 0 & 0 & 0.1 & 0.4 \\ 0 & 0 & 0 & 0.375 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 0.1 & 0.085 \\ 0 & 1 & 0.8 & 0.5 \\ 0 & 0 & 0.1 & 0.04 \\ 0 & 0 & 0 & 0.375 \end{pmatrix}.$$

Hence, we obtain by (3.3) and since $\sum_{l \in \text{An}(i)} b_{li} = 1$ for all $i = 1, \dots, 4$,

$$\chi(2, 3) = 0.8, \quad \chi(3, 4) = 0.625, \quad \text{and} \quad \chi(2, 4) = 0.5.$$

Hence, $\chi(2, 3)\chi(3, 4) = \chi(2, 4)$. However, $b_{23} \frac{b_{34}}{b_{33}} = 0.8 \cdot 0.4 < 0.5 = b_{24}$, implying that the path $[2 \rightarrow 3 \rightarrow 4]$ is not max-weighted. \square

4 Identifiability from the tail dependence coefficients

The goal of this section is to investigate how far the matrix of tail dependence coefficients $(\chi(i, j))_{i, j \in V}$ of a recursive ML model on a DAG with regularly varying noise Z_1, \dots, Z_d determines the ML coefficient matrix B and, hence, the reachability matrix $R = \text{sgn}(B)$.

Definition 4.1. We call two recursive ML models \mathbf{X}_1 and \mathbf{X}_2 χ -equivalent, if their tail dependence coefficient matrices agree, i.e., if $\chi_1 = \chi_2$. \square

The definition has its analogue in the classical framework of structural equation models with linear functions and Gaussian noise. For instance, it is shown in Heckerman and Geiger [9] that for all graphs in the same Markov equivalence class there exists a structural equation model that leads to the same distribution \mathbf{X} . Extensions and ramifications can be found in Peters [14].

The following example investigates the relation between the χ -equivalence of two recursive ML models and the Markov equivalence of their underlying DAGs.

Example 4.2. (1) Consider the following Markov equivalent DAGs \mathcal{D}_1 (left) and \mathcal{D}_2 (right).



We show that one cannot find two ML coefficient matrices B_1 and B_2 such that the resulting recursive ML models \mathbf{X}_1 and \mathbf{X}_2 are χ -equivalent. We call the tail dependence coefficient matrix of the first model χ_1 and of the second model χ_2 , respectively. By Proposition 5.4(a) below, we have $\chi_1(1,3) = \chi_1(1,2)\chi_1(2,3)$, implying by Lemma 3.6(a) that $\chi_1(1,3) < \chi_1(1,2) \wedge \chi_1(2,3)$. Concerning the second model, we have by Theorem 5.2 that $\chi_2(1,3) = \chi_2(1,2) \wedge \chi_2(2,3)$. Hence, \mathcal{D}_1 and \mathcal{D}_2 do not correspond to χ -equivalent ML models.

(2) Conversely, we demonstrate that there exist χ -equivalent recursive ML models \mathbf{X}_1 on \mathcal{D}_1 (left) and \mathbf{X}_2 on \mathcal{D}_2 (right) such that \mathcal{D}_1 and \mathcal{D}_2 are not Markov equivalent.



Let χ_1 denote the tail dependence coefficient matrix of \mathbf{X}_1 and suppose $\chi_1(1,2) \leq \chi_1(1,3)$. We denote the ML coefficient matrix of \mathbf{X}_2 by B_2 and choose B_2 such that $\overline{B}_2(2,1) := \chi_1(1,2)$ and $\overline{B}_2(2,3) := \chi_1(2,3)$. An easy calculation shows that $\chi_1 = \chi_2$, meaning that \mathbf{X}_1 and \mathbf{X}_2 are χ -equivalent. However, the undirected versions of \mathcal{D}_1 and \mathcal{D}_2 do not agree, implying that \mathcal{D}_1 and \mathcal{D}_2 are not Markov equivalent. \square

For $d \in \mathbb{N}$ let M_d denote the set of all ML coefficient matrices B corresponding to a DAG with d nodes. We consider the function

$$f_d : M_d \rightarrow \mathbb{R}_+^{d \times d}, \quad B \mapsto \chi_B, \quad (4.1)$$

where χ_B denotes the tail dependence matrix corresponding to B . As observed in (3.3), the matrix χ_B is uniquely determined by \overline{B} and the map $B \mapsto \overline{B}$ is not injective, also f_d is not injective. The following example shows that $\chi_{B_1} = \chi_{B_2}$ does not even imply $\overline{B}_1 = \overline{B}_2$; i.e., that the function f_d is not injective when restricted to matrices with column sum equal to one.

Example 4.3. For $b \in (0,1)$ consider the ML coefficient matrices

$$B_1 = \begin{pmatrix} 1 & b \\ 0 & 1-b \end{pmatrix} \quad \text{and} \quad B_2 = \begin{pmatrix} 1-b & 0 \\ b & 1 \end{pmatrix}.$$

We obtain, since $\chi(1,2) = b_{12}$, $\chi(2,1) = b_{21}$, and $\chi(1,1) = \chi(2,2) = 1$,

$$\chi_{B_1} = \chi_{B_2} = \begin{pmatrix} 1 & b \\ b & 1 \end{pmatrix}.$$

The point here is that the tail dependence coefficient $\chi(1,2) = \chi(2,1)$ is symmetric and, hence, cannot give the direction of the edge between 1 and 2. \square

From this example we understand that the labels of the nodes are relevant. For what follows we need the definition of a topological order of the nodes of a DAG.

Definition 4.4. Let $\mathcal{D} = (V, E)$ be a DAG with d nodes. We call $\tau : V \rightarrow \{1, \dots, d\}$, defined by $i \mapsto \tau(i)$, a *topological order*, if for every edge $(i, j) \in E$ we have $\tau(i) < \tau(j)$. \square

Note that τ is a topological order if and only if for every path $[i \Rightarrow j] \subseteq \mathcal{D}$ we have $\tau(i) < \tau(j)$. Hence, recalling from (1.3) that different DAGs corresponding to the same ML coefficient matrix B have the same reachability matrix $\text{sgn}(B)$, we conclude that τ is a topological order for every DAG corresponding to B , if it is a topological order for one representative.

It is not hard to see that for a given DAG a topological order is in general not unique. On the other hand, by Diestel [5], Appendix A, there exists a topological order for every DAG.

The following result shows that the function f_d from (4.1), restricted to matrices B with column sum equal to one that correspond to DAGs with a fixed topological order τ , is injective. This means that the matrix \overline{B} can be identified from $(\chi(i, j))_{i, j \in V}$ if a topological order τ of \mathcal{D} is given and known. Equivalently, the exponent measure μ_* from Remark 2.2 that is uniquely determined by \overline{B} can be identified from $(\chi(i, j))_{i, j \in V}$ and τ . In particular, we find the reachability matrix $\text{sgn}(\overline{B})$ of \mathcal{D} .

Theorem 4.5. *Let B_1 and B_2 be ML coefficient matrices and τ be a topological order for the DAGs corresponding to B_1 and B_2 . Then $\chi_{B_1} = \chi_{B_2}$ implies $\overline{B}_1 = \overline{B}_2$.*

Proof. We apply a double induction and show that given a topological order τ , \overline{b}_{ji} is uniquely determined by χ for $i, j \in V$. The outer induction runs over $\tau(i)$. For $\tau(i) = 1$, i is an initial node and, hence, $\overline{b}_{ji} = \mathbf{1}_{\{i\}}(j)$. The induction hypothesis states that \overline{b}_{ji} is known for all $j \in V$ and $\tau(i) \leq n$. In the induction step we use the induction hypothesis to determine \overline{b}_{ji} for all $j \in V$ and $\tau(i) = n + 1$.

Here we fix i and start the inner induction that runs over $\tau(j)$. For $\tau(j) = 1$, we have $\overline{b}_{ji} = \chi(j, i)$ by Proposition 3.6(c), since $\tau(j) = 1$ implies $j \in V_0$. The induction hypothesis of the inner induction says that we know \overline{b}_{ji} for $\tau(j) \leq m$ and $\tau(i) \leq n + 1$. In the induction step of the inner induction we determine \overline{b}_{ji} for $\tau(j) = m + 1$ and $\tau(i) = n + 1$, where $m \leq n$. For $m > n$, $j \in V \setminus \text{An}(i)$ and, hence, $b_{ji} = 0$. By (3.3), we have

$$\chi(j, i) = \sum_{k \in \text{An}(j) \cap \text{An}(i)} \overline{b}_{kj} \wedge \overline{b}_{ki} = \overline{b}_{ji} \wedge \overline{b}_{jj} + \sum_{k \in \text{an}(j) \cap \text{an}(i)} \overline{b}_{kj} \wedge \overline{b}_{ki},$$

Since $\overline{b}_{ji} \leq \overline{b}_{jj}$ by Remark 3.3 and Lemma 2.4, we obtain

$$\overline{b}_{ji} = \chi(j, i) - \sum_{k \in \text{an}(j) \cap \text{an}(i)} \overline{b}_{kj} \wedge \overline{b}_{ki} = \chi(j, i) - \sum_{k: \tau(k) < \tau(j) \wedge \tau(i)} \overline{b}_{kj} \wedge \overline{b}_{ki}. \quad (4.2)$$

Now we exploit the induction hypotheses of the outer and inner induction and note that the rhs of (4.2) is known. Hence, we can find \overline{b}_{ji} and complete the induction step of the inner induction. This is exactly what was needed for the induction step of the outer induction. Therefore, also the outer induction is finished and the proof is complete. \square

In what follows we characterize the set of initial nodes V_0 of a DAG \mathcal{D} by means of the tail dependence coefficient matrix χ . Our motivation to do this is caused by the important role of the initial nodes in the identification of \overline{B} from χ . In Theorem 5.6 below we will see that V_0 and χ together determine \overline{B} for every max-weighted model.

Proposition 4.6. *Let \mathbf{X} be a recursive ML model on a DAG \mathcal{D} with ML coefficient matrix $B = (b_{ij})_{d \times d}$ and tail dependence matrix $\chi = (\chi(i, j))_{d \times d}$. Let V_0 denote the set of initial nodes of \mathcal{D} .*

(a) *For all $i, j \in V_0$ we have $\chi(i, j) = 0$.*

(b) Let $W \subseteq V$ such that $\chi(i, j) = 0$ for all $i, j \in W$. Then $|W| \leq |V_0|$.

Proof. (a) For all $i, j \in V_0$ we have $\text{An}(i) \cap \text{An}(j) = \emptyset$, then apply Theorem 3.4.

(b) Assume $|W| > |V_0|$. Then there are $i_1, i_2 \in W$ and $j \in V_0$ with $j \in \text{An}(i_1)$ and $j \in \text{An}(i_2)$. In particular, $\text{An}(i_1) \cap \text{An}(i_2) \neq \emptyset$. This contradicts by Theorem 3.4 the assumption $\chi(i_1, i_2) = 0$. Therefore, $|W| \leq |V_0|$. \square

The tail dependence coefficient matrix χ induces an undirected graph on V with an edge $j \rightarrow i$, whenever the tail dependence coefficient $\chi(i, j) \neq 0$. We consider here the dual to the graph induced by χ . We also define the notion of a clique in this dual graph.

Definition 4.7. Let $W \subseteq V$.

(a) We call W a χ -clique of \mathcal{D} if it is a clique in the dual graph induced by χ .

(b) We call W a *maximum* χ -clique of \mathcal{D} if it is a maximum clique in the dual graph induced by χ . \square

Remark 4.8. (a) Proposition 4.6 says that V_0 is a maximum χ -clique.

(b) Let W be a maximum χ -clique. We know from Theorem 3.4 that $\chi(i, j) = 0$ if and only if $\text{An}(i) \cap \text{An}(j) \neq \emptyset$. Hence, $|\text{An}(i) \cap V_0| = 1$ for all $i \in W$. \square

5 Recursive ML max-weighted models on DAGs

In this section we investigate in detail the recursive ML max-weighted models introduced in Definition 2.8. We extensively use the property that in the corresponding DAG every path is max-weighted.

Proposition 5.1. Let \mathbf{X} be a recursive ML max-weighted model on a DAG \mathcal{D} and $i, j \in V$.

(a) For $i \in \text{An}(j)$ we have $\chi(i, j) = \frac{\bar{b}_{ij}}{\bar{b}_{ii}}$.

(b) For $i \in \text{pa}(j)$ we have $\chi(i, j) = c_i^j \frac{\sum_{l=1}^d b_{li}^\alpha}{\sum_{l=1}^d b_{lj}^\alpha}$.

(c) For $j \in \text{An}(i)$ with path $[j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$ we have $\chi(i, j) = \chi(k_0, k_1) \cdots \chi(k_{n-1}, k_n)$.

Proof. (a) is a consequence of Proposition 3.6(b),

(b) follows from (a) and (1.3), since $\chi(i, j) = \frac{\bar{b}_{ij}}{\bar{b}_{ii}} = \frac{b_{ij}}{b_{ii}} \frac{\sum_{l=1}^d b_{li}^\alpha}{\sum_{l=1}^d b_{lj}^\alpha} = c_i^j \frac{\sum_{l=1}^d b_{li}^\alpha}{\sum_{l=1}^d b_{lj}^\alpha}$, and

(c) is a consequence of (a) since $\frac{\bar{b}_{lk} \bar{b}_{km}}{\bar{b}_{kk}} = \bar{b}_{lm}$ for all $k \in \text{De}(l) \cap \text{An}(m)$ in a max-weighted model. \square

The following result shows that every tail dependence coefficient in a max-weighted model is a linear combination of the minima of tail dependence coefficients along paths.

Theorem 5.2. Let \mathbf{X} be a recursive ML max-weighted model on a DAG \mathcal{D} and $i, j \in V$. Then we have

$$\chi(i, j) = \sum_{k \in \text{An}(i) \cap \text{An}(j)} (\chi(k, i) \wedge \chi(k, j)) \lambda_k, \quad (5.1)$$

with $\lambda_k = 1 - \sum_{l \in \text{de}(k) \cap \text{An}(i) \cap \text{An}(j)} \lambda_l$.

Proof. First recall that every DAG admits a topological order, such that the sum in (5.1) is well-defined. Let $i, j \in V$ and $k \in \text{An}(i) \cap \text{An}(j)$. Since in a max-weighted model $\bar{b}_{li} = \frac{\bar{b}_{lk}\bar{b}_{ki}}{\bar{b}_{kk}}$ for all $l \in \text{An}(k)$ (cf. (2.8)), and $\frac{\bar{b}_{ki}}{\bar{b}_{kk}} < 1$ by Lemma 2.4, we have

$$\begin{aligned}\chi(k, i) \wedge \chi(k, j) &= \left(\sum_{l \in \text{An}(k)} \bar{b}_{lk} \wedge \bar{b}_{li} \right) \wedge \left(\sum_{l \in \text{An}(k)} \bar{b}_{lk} \wedge \bar{b}_{lj} \right) \\ &= \left(\sum_{l \in \text{An}(k)} \bar{b}_{li} \right) \wedge \left(\sum_{l \in \text{An}(k)} \bar{b}_{lj} \right) \\ &= (\bar{b}_{ki} \wedge \bar{b}_{kj}) \sum_{l \in \text{An}(k)} \frac{\bar{b}_{lk}}{\bar{b}_{kk}} \\ &= \sum_{l \in \text{An}(k)} \bar{b}_{li} \wedge \bar{b}_{lj}.\end{aligned}$$

Hence,

$$\sum_{k \in \text{An}(i) \cap \text{An}(j)} (\chi(k, i) \wedge \chi(k, j)) \lambda_k = \sum_{k \in \text{An}(i) \cap \text{An}(j)} \lambda_k \sum_{l \in \text{An}(k)} \bar{b}_{li} \wedge \bar{b}_{lj}.$$

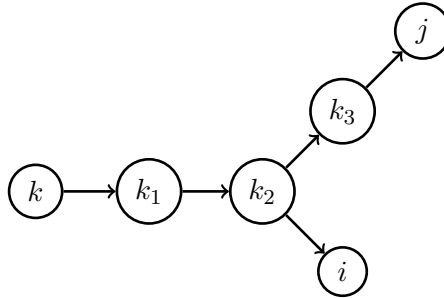
Interchanging the sums yields

$$\begin{aligned}\sum_{k \in \text{An}(i) \cap \text{An}(j)} (\chi(k, i) \wedge \chi(k, j)) \lambda_k &= \sum_{l \in \text{An}(i) \cap \text{An}(j)} (\bar{b}_{li} \wedge \bar{b}_{lj}) \sum_{k \in \text{De}(l) \cap \text{An}(i) \cap \text{An}(j)} \lambda_k \\ &= \sum_{l \in \text{An}(i) \cap \text{An}(j)} (\bar{b}_{li} \wedge \bar{b}_{lj}) \left(\lambda_l + \sum_{k \in \text{de}(l) \cap \text{An}(i) \cap \text{An}(j)} \lambda_k \right) \\ &= \sum_{l \in \text{An}(i) \cap \text{An}(j)} \bar{b}_{li} \wedge \bar{b}_{lj} = \chi(i, j).\end{aligned}$$

□

We want to illustrate this result for an example.

Example 5.3. Consider the following DAG \mathcal{D} :



We find from Theorem 5.2 that

$$\chi(i, j) = (\chi(k, i) \wedge \chi(k, j)) \lambda_k + (\chi(k_1, i) \wedge \chi(k_1, j)) \lambda_{k_1} + (\chi(k_2, i) \wedge \chi(k_2, j)) \lambda_{k_2}$$

with $\lambda_k = 1 - \lambda_{k_1} - \lambda_{k_2}$, $\lambda_{k_1} = 1 - \lambda_{k_2}$, and $\lambda_{k_2} = 1$. □

Proposition 5.4. Let \mathbf{X} be a recursive ML max-weighted model on a DAG \mathcal{D} and $i, j \in V$. Let V_0 denote the set of initial nodes of \mathcal{D} .

(a) Let $k \in V$. Then $k \in \text{An}(i)$ if and only if $\chi(k, i) > 0$ and for all $j \in \text{An}(i) \cap \text{An}(k) \cap V_0$, we have $\chi(j, i) = \chi(j, k)\chi(k, i)$.

(b) Let $k_0, \dots, k_n \in V$. There is a path $[k_0 \rightarrow \dots \rightarrow k_n]$ if and only if $\chi(k_m, k_{m+1}) > 0$ for $m = 0, \dots, n-1$ and for all $j \in \text{An}(k_0) \cap V_0$ and $0 \leq m_1 \leq m_2 \leq n$ we have $\chi(j, k_{m_2}) = \chi(j, k_{m_1})\chi(k_{m_1}, k_{m_2})$.

Proof. (a) Assume that $k \in \text{An}(i)$. Thus, we have by Proposition 5.1(a) that $\chi(k, i) = \frac{\bar{b}_{ki}}{\bar{b}_{kk}}$. Let $j \in \text{An}(i) \cap \text{An}(k) \cap V_0$ or, equivalently, $j \in V_0$ such that $\chi(j, k) > 0$ and $\chi(j, i) > 0$. Hence, $\chi(j, k) = \frac{\bar{b}_{jk}}{\bar{b}_{jj}} = \bar{b}_{jk}$ and $\chi(j, i) = \frac{\bar{b}_{ji}}{\bar{b}_{jj}} = \bar{b}_{ji}$ since $\bar{b}_{jj} = 1$ for $j \in V_0$. Since every path from j to i passing through k is max-weighted, we have by Lemma 2.10(a) of [8] that $b_{ji} = \frac{b_{jk}b_{ki}}{b_{kk}}$, implying $\bar{b}_{ji} = \frac{\bar{b}_{jk}\bar{b}_{ki}}{\bar{b}_{kk}}$. Thus,

$$\chi(j, i) = \bar{b}_{ji} = \frac{\bar{b}_{jk}\bar{b}_{ki}}{\bar{b}_{kk}} = \chi(j, k)\chi(k, i).$$

Now assume that $\chi(j, i) = \chi(j, k)\chi(k, i)$ for all $j \in \text{An}(i) \cap \text{An}(k) \cap V_0$ or, equivalently, for all $j \in V_0$ such that $\chi(j, k) > 0$ and $\chi(j, i) > 0$. By Proposition 5.1(a) we have

$$\chi(k, i) = \frac{\chi(j, i)}{\chi(j, k)} = \frac{\bar{b}_{ji}}{\bar{b}_{jk}} \text{ for all } j \in \text{An}(i) \cap \text{An}(k) \cap V_0. \quad (5.2)$$

Combining (3.3) and (5.2), we obtain for $j \in \text{An}(i) \cap \text{An}(j) \cap V_0$,

$$\frac{\bar{b}_{ji}}{\bar{b}_{jk}} = \chi(k, i) = \sum_{l \in \text{An}(i) \cap \text{An}(k)} \bar{b}_{li} \wedge \bar{b}_{lk},$$

equivalently,

$$\bar{b}_{ji} = \bar{b}_{ji} \sum_{l \in \text{An}(i) \cap \text{An}(k)} \frac{\bar{b}_{li}\bar{b}_{jk}}{\bar{b}_{ji}} \wedge \frac{\bar{b}_{lk}\bar{b}_{jk}}{\bar{b}_{ji}}. \quad (5.3)$$

Next we show that $\frac{\bar{b}_{li}\bar{b}_{jk}}{\bar{b}_{ji}} = \bar{b}_{lk}$ for all $l \in \text{An}(i) \cap \text{An}(k)$. Since $\chi(k, i) > 0$, $\text{An}(i) \cap \text{An}(k) \neq \emptyset$ by Theorem 3.4. Let $\tilde{j} \in V_0$ such that $\tilde{j} \in \text{An}(l)$, and recall that $\chi(k, i) = \frac{\bar{b}_{ji}}{\bar{b}_{jk}} = \frac{\bar{b}_{ji}}{\bar{b}_{\tilde{j}k}}$, since $\tilde{j} \in \text{An}(i) \cap \text{An}(k) \cap V_0$. Thus, using $\bar{b}_{\tilde{j}k} = \frac{\bar{b}_{\tilde{j}l}\bar{b}_{lk}}{\bar{b}_{ll}}$ and $\bar{b}_{\tilde{j}i} = \frac{\bar{b}_{\tilde{j}l}\bar{b}_{li}}{\bar{b}_{ll}}$ since all paths are max-weighted (cf. Lemma 2.10(a) of [8]), we have

$$\frac{\bar{b}_{li}\bar{b}_{jk}}{\bar{b}_{ji}} = \frac{\bar{b}_{li}\bar{b}_{\tilde{j}k}}{\bar{b}_{\tilde{j}i}} = \frac{\bar{b}_{li}\bar{b}_{\tilde{j}l}\bar{b}_{lk}\bar{b}_{ll}}{\bar{b}_{ll}\bar{b}_{\tilde{j}l}\bar{b}_{li}} = \bar{b}_{lk}. \quad (5.4)$$

Using (5.4) and recalling from (5.2) that $\frac{\bar{b}_{ji}}{\bar{b}_{jk}} = \chi(k, i) \leq 1$, (5.3) becomes

$$\bar{b}_{ji} = \bar{b}_{ji} \sum_{l \in \text{An}(i) \cap \text{An}(k)} \bar{b}_{lk} \wedge \frac{\bar{b}_{lk}\bar{b}_{jk}}{\bar{b}_{ji}} = \bar{b}_{ji} \sum_{l \in \text{An}(i) \cap \text{An}(k)} \bar{b}_{lk}.$$

Since $\bar{b}_{ji} > 0$ and $\sum_{l \in \text{An}(k)} \bar{b}_{lk} = 1$, this implies $\bar{b}_{lk} = 0$ for $l \in \text{An}(i) \setminus \text{An}(k)$, equivalently $\text{An}(k) \subseteq \text{An}(i)$, implying $k \in \text{An}(i)$.

(b) is a consequence of (a). \square

Remark 5.5. Prop. 5.4(a) relates to Prop. 5.2 of [12]. Indeed $\chi(j, i) = \chi(j, k)\chi(k, i)$ is equivalent to the (ji) -th component of the inverse of the 3×3 -matrix $(\chi(m, n))_{m, n=i, j, k}$ being 0. For multivariate normal distributions this corresponds to the (ji) -th component of the concentration matrix being 0. \square

Theorem 5.6. *Let \mathbf{X} be a recursive ML max-weighted model on a DAG \mathcal{D} . Then the matrix \overline{B} is identifiable from the tail dependence matrix χ and the initial nodes V_0 .*

Proof. First we identify the reachability matrix of \mathcal{D} from the tail dependence matrix χ and V_0 . We know from Proposition 5.4(a) that for $i, k \in V$, $k \in \text{An}(i)$ if and only if $\chi(k, i) > 0$ and $\chi(j, i) = \chi(j, k)\chi(k, i)$ for all nodes $j \in V_0$ with $\chi(j, i) > 0$ and $\chi(j, k) > 0$. Thus we find for every node its ancestors and, hence, the reachability matrix of \mathcal{D} . In order to find \overline{B} , recall from (1.3) that $\overline{b}_{ji} = 0$ for all $i \in V$ and $j \in V \setminus \text{An}(i)$. For $i \in V$ and $j \in \text{An}(i)$ we have by Proposition 5.1(a) that $\overline{b}_{ji} = \overline{b}_{jj} \chi(i, j)$. Hence, for all $i \in V$, as $\sum_{l \in \text{An}(i)} \overline{b}_{li} = 1$,

$$\overline{b}_{ii} = 1 - \sum_{j \in \text{an}(i)} \overline{b}_{ji} = 1 - \sum_{j \in \text{an}(i)} \overline{b}_{jj} \chi(i, j).$$

Observe for all $j \in V_0$ that $\overline{b}_{jj} = 1$. Consequently, we may proceed iteratively starting with the nodes with minimal number of ancestors to determine \overline{B} . \square

Remark 5.7. Theorem 5.6 is a refinement of Theorem 4.5 for the max-weighted model. Instead of knowing the whole topological order of the corresponding DAG, it suffices to know the initial nodes V_0 in order to identify \overline{B} from χ in the max-weighted model. \square

In order to determine the reachability matrix of \mathcal{D} we do not have to verify for all pairs of nodes $i, k \in V$ and all $j \in V_0$ with $\chi(j, k)\chi(j, i) > 0$, whether $\chi(j, k)\chi(k, i) = \chi(j, i)$. The following result shows that ancestral relations in the DAG \mathcal{D} imply certain monotonicities in the tail dependence matrix χ that can be used in its identification.

Proposition 5.8. *Let \mathbf{X} be a recursive ML max-weighted model on a DAG \mathcal{D} .*

- (a) *Let $i, k \in V$. Then $\text{An}(k) \cap \text{An}(i) = \emptyset$ if and only if $\chi(j, k)\chi(j, i) = 0$ for all $j \in V_0$.*
- (b) *Let $i, k \in V$ such that $\chi(j, k)\chi(j, i) > 0$ for some $j \in V_0$. If $\chi(j, k) < \chi(j, i)$ for some $j \in \text{An}(k) \cap \text{An}(i) \cap V_0$, then $k \in V \setminus \text{An}(i)$.*

Proof. (a) Observe that $\text{An}(k) \cap \text{An}(i) = \emptyset$ if and only if $\text{An}(k) \cap \text{An}(i) \cap V_0 = \emptyset$. Recall from Lemma 5.4 that $i, k \in \text{De}(j)$ for some $j \in V_0$ if and only if $\chi(j, k)\chi(j, i) > 0$. Thus $\text{An}(k) \cap \text{An}(i) = \emptyset$ if and only if $\chi(j, k)\chi(j, i) = 0$ for all $j \in V_0$.

(b) Recall from Lemma 5.4(a) that $\text{An}(i) \cap \text{An}(k) \cap V_0 \neq \emptyset$. Assume $\chi(j, k) < \chi(j, i)$ for some $j \in V_0$, but $k \in \text{An}(i)$. Since by Lemma 5.4, $\chi(j, i) = \overline{b}_{ji}$ and $\chi(j, k) = \overline{b}_{jk}$, and all paths are max-weighted, we obtain from (2.4) that

$$\chi(j, i) = \overline{b}_{ji} = \frac{\overline{b}_{jk}\overline{b}_{ki}}{\overline{b}_{kk}} = \chi(j, k) \frac{\overline{b}_{ki}}{\overline{b}_{kk}} \leq \chi(j, k),$$

which is a contradiction to $\chi(j, k) < \chi(j, i)$. \square

The theory developed so far provides a recipe to find all matrices \overline{B} corresponding to a given tail dependence matrix χ . First, calculate all maximum cliques as in Definition 4.7(b). If there is only one maximum clique, we know from Proposition 4.6 that this is the set of initial nodes V_0 . If there are several maximum cliques, χ might correspond to different ML models on DAGs with different sets of initial nodes. To find all of them, we proceed as described in the proof of Theorem 5.6 and construct a reachability matrix from χ and every maximum clique. Note that this is always possible and does not require that χ is indeed the tail dependence matrix of a recursive ML model on a DAG. Therefore, we recall from Theorem 5.2 that χ is uniquely determined by the entries $\chi(i, j)$ for $i \in \text{An}(j)$. Hence, we check whether χ is consistent with the reachability matrices constructed above. If this is the case, we find the matrix \overline{B} as described in the proof of Theorem 5.6. Otherwise, the maximum clique is not the set of initial nodes for a recursive ML model on any DAG corresponding to χ .

The following theorem shows the link between the initial nodes of the DAGs corresponding to different ML models with the same tail dependence matrix χ .

Theorem 5.9. *Let B, B' be two ML coefficient matrices corresponding to a max-weighted model on a DAG \mathcal{D} with the initial nodes V_0 and \tilde{V}_0 , respectively, that share the same tail dependence matrix χ . Then*

$$V'_0 \subseteq V_0 \cup V_\infty,$$

where V_∞ are the terminal nodes for B .

Proof. Assume that there is some $k \in V'_0$ s.t. $k \notin V_0 \cup V_\infty$. By Lemma 5.4, there are nodes $j \in \text{an}(k)$ and $i \in \text{de}(k)$ with

$$\chi(j, i) = \chi(j, k)\chi(k, i). \quad (5.5)$$

Since $\chi(j, k) < 1$ and $\chi(k, i) < 1$ by Lemma 3.6(a), (5.5) implies

$$\chi(j, i) < \chi(j, k) \wedge \chi(k, i). \quad (5.6)$$

Since $j \in \text{An}(k) \cap \text{An}(i)$, $\chi(j, k) > 0$ and $\chi(k, i) > 0$ by Theorem 3.4. Therefore, exploiting that χ is the tail dependence matrix for B' , $k \in \text{An}'(j) \cap \text{An}'(i) \cap \tilde{V}_0$. Hence,

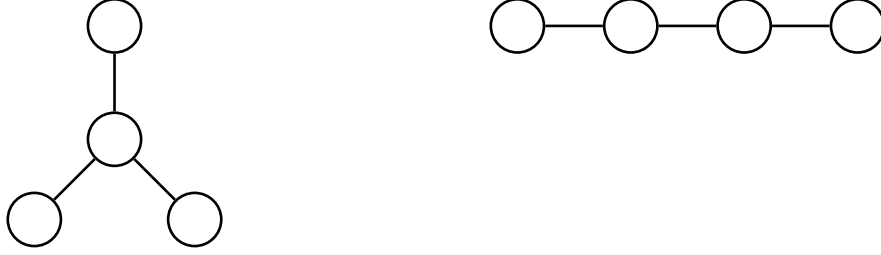
$$\chi(j, i) = \sum_{l \in \text{An}'(j) \cap \text{An}'(i)} \overline{b}'_{lj} \wedge \overline{b}'_{li} \geq \overline{b}'_{kj} \wedge \overline{b}'_{ki} = \chi(j, k) \wedge \chi(k, i),$$

which contradicts (5.6). \square

The following example illustrates the identifiability problem, when only the tail dependence matrix is given.

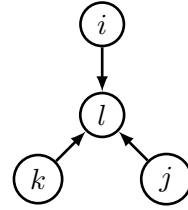
Example 5.10. [Polytree with $d = 4$ nodes]

The corresponding (undirected) tree is one of those:

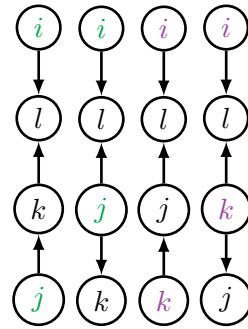


We investigate the identifiability problem, where we assume that the tail dependence coefficient matrix χ is known. Let $V = \{i, j, k, l\}$. We classify the possible situations with respect to the initial nodes V_0 .

(1) Assume that $\chi(i, j) = \chi(i, k) = \chi(j, k) = 0$ and all others > 0 . Then $V_0 = \{X_i, X_j, X_k\}$ such that the DAG is identifiable and given by the one on the right hand side.



(2) Assume that $\chi(i, j) = \chi(i, k) = 0$ and all others > 0 . Then either $V_0^1 = \{X_i, X_j\}$ or $V_0^2 = \{X_i, X_k\}$. We apply Proposition 5.4 and make all possible comparisons: If $\chi(j, k)\chi(k, l) = \chi(j, l)$, then we obtain the first DAG on the right hand side, and if $\chi(k, j)\chi(j, l) = \chi(k, l)$, we obtain the third DAG. If neither of this is true we must have either the second or the fourth DAG. We can, however, apply Theorem 5.2 in this situation. It implies for the second DAG that $\chi(k, l) = \chi(j, k) \wedge \chi(j, l)$ and for the fourth DAG that $\chi(l, j) = \chi(l, k) \wedge \chi(j, k)$. Thus, if $\chi(j, k) = \chi(j, k) \wedge \chi(j, l) \wedge \chi(l, k)$, then we cannot distinguish between the two DAGs, in all other cases we can.

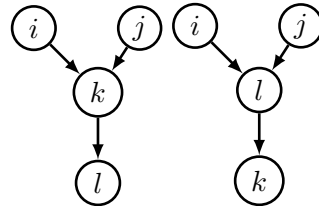


(3) Assume that $\chi(i, j) = 0$ and all others > 0 . Then $V_0 = \{i, j\}$. We apply Proposition 5.4 and make the possible comparisons:

$$\chi(i, k)\chi(k, l) = \chi(i, l) \text{ and } \chi(i, l)\chi(l, k) = \chi(i, k)$$

$$\chi(j, k)\chi(k, l) = \chi(j, l) \text{ and } \chi(j, l)\chi(l, k) = \chi(j, k)$$

They can only hold pairwise, either the first in the first and second line, or the second in both lines. Thus, we find one of the DAGs on the right hand side.



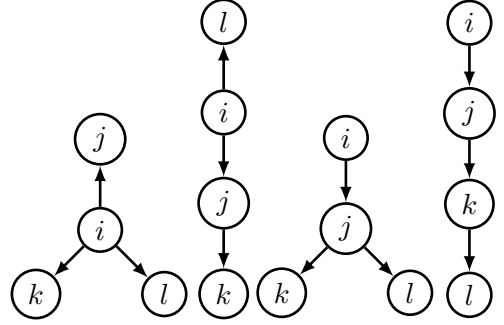
(4) Assume that all $\chi(i, j) > 0$. We apply Proposition 5.4 and check how often we have $\chi(i, j)\chi(j, k) = \chi(i, k)$ for every tripel (i, j, k) .

- If this never holds, then we must have the first DAG on the right hand side, and cannot decide with this criterion, which node is the center node. Thus, we have 4 possible DAGs. We can again apply Theorem 5.2 in this situation. It implies that $\chi(j, k) = \chi(i, j) \wedge \chi(i, k)$, $\chi(j, l) = \chi(i, j) \wedge \chi(i, l)$, and $\chi(k, l) = \chi(i, k) \wedge \chi(i, l)$. In order to identify the nodes, we order all $\chi(\cdot, \cdot)$, and assume that $\chi(i, j)$ is minimal, and that $\chi(i, j) = \chi(j, k) = \chi(j, l)$, then i must be the center node.

-If only $\chi(i, j)\chi(j, k) = \chi(i, k)$ holds (or any other combination of nodes), then we must have the second DAG.

- If $\chi(i, j)\chi(j, k) = \chi(i, k)$ and $\chi(i, j)\chi(j, l) = \chi(i, l)$ (or another combination of nodes), then we have the third DAG.

-If $\chi(i, j)\chi(j, k) = \chi(i, k)$, $\chi(i, j)\chi(j, l) = \chi(i, l)$ and $\chi(j, k)\chi(k, l) = \chi(j, l)$, then we have the fourth DAG. However, since χ is symmetric, also the DAG with all directions inverted, is also a possible DAG to this situation (cf. Theorem 5.9).



By applying Proposition 5.4 and Theorem 5.2, all polytrees with 4 nodes are identifiable with the exception of the last one, where we cannot distinguish between this or the one with reversed directions (cf. Theorem 5.9). \square

6 Statistical estimation of a recursive ML model

We assume the distributional situation as in Section 2; i.e., the data $\mathbf{X} = (X_1, \dots, X_d)$ is (marginally) regularly varying with index $\alpha > 0$ and the vector \mathbf{X} has (after standardization of the marginals) exponent measure $\mu_*([\mathbf{0}, \mathbf{x}]^c) = \sum_{k=1}^d \bigvee_{i \in \text{De}(k)} \frac{b_{ki}^\alpha}{\sum_{l=1}^d b_{li}^\alpha} x_i^{-1}$ (cf. Remark 2.2).

The goal of this section is to provide estimators for the tail-dependence matrix $\chi = (\chi(i, j))_{i,j=1,\dots,d}$ and to derive its asymptotic properties.

6.1 A consistent and asymptotically normal estimator for the matrix χ

Denote by F_i the marginal distribution function of X_i for $i = 1, \dots, d$. In (3.1) we defined the tail dependence coefficient between X_i and X_j by

$$\chi(i, j) := \lim_{u \rightarrow 1} \frac{\mathbb{P}(F_i(X_i) > u, F_j(X_j) > u)}{\mathbb{P}(F_i(X_i) > u)}.$$

Assume we observe $(X_{i,1}, \dots, X_{i,n})$ for $i = 1, \dots, d$ and denote by $X_{i,1:n} < X_{i,2:n} < \dots < X_{i,n:n}$ the order statistics of the i -th component of the observations. A natural non-parametric estimator of χ is given by choosing $x := \frac{n-k}{n}$ for some sufficiently large order statistics and defining for

$i, j \in V$ the empirical counterpart of $\chi(i, j)$ as

$$\begin{aligned}\chi^{n,k}(i, j) &:= \frac{\sum_{m=1}^n \mathbf{1}_{\{R(X_{i,m}) > n-k \text{ and } R(X_{j,m}) > n-k\}}}{\sum_{m=1}^n \mathbf{1}_{\{R(X_{i,m}) > n-k\}}} \\ &= \frac{1}{k} \sum_{m=1}^n \mathbf{1}_{\{R(X_{i,m}) > n-k \text{ and } R(X_{j,m}) > n-k\}} \\ &= \frac{1}{k} \sum_{m=1}^n \mathbf{1}_{\{X_{i,m} > X_{i,n-k:n} \text{ and } X_{j,m} > X_{j,n-k:n}\}},\end{aligned}$$

where $R(X_{i,l}) := \sum_{m=1}^n \mathbf{1}_{\{X_{i,m} \leq X_{i,l}\}}$

denotes the rank of $X_{i,l}$ among $(X_{i,1}, \dots, X_{i,n})$. Since the estimator $\chi^{n,k}$ only relies on ranks, it is independent of the marginal distribution F_1, \dots, F_d . For the proof of asymptotic normality of $\chi^{n,k}$ we study the function $V^{n,k}$ defined by

$$V^{n,k}(x_1, \dots, x_d) := \frac{1}{k} \sum_{m=1}^n \mathbf{1}_{\{F_1(X_{1,m}) > 1-kx_1/n \text{ and } \dots \text{ and } F_d(X_{d,m}) > 1-kx_d/n\}} \quad (6.1)$$

and write $V_{ij}^{n,k} := V^{n,k}(e_i + e_j)$. $V^{n,k}$ can be understood as an estimator of χ for known marginal distributions F_1, \dots, F_d . Note that

$$\chi^{n,k}(i, j) = V^{n,k}\left(\frac{n}{k}(1 - F_i(X_{i,n-k:n}))e_i + \frac{n}{k}(1 - F_j(X_{j,n-k:n}))e_j\right). \quad (6.2)$$

We obtain the following theorem.

Theorem 6.1. *Let $(i_1, j_1), \dots, (i_h, j_h)$ be arbitrary pairs of components of \mathbf{X} . Suppose that for $k = 1, \dots, h$ and $\beta > 0$,*

$$x \mathbb{P}\left(F_{i_k}(X_{i_k}) > 1 - \frac{1}{x}, F_{j_k}(X_{j_k}) > 1 - \frac{1}{x}\right) = \chi(i_k, j_k) + O(x^{-\beta}).$$

Let $k = k(n) \rightarrow \infty$ as $n \rightarrow \infty$ and $k = o(n^{2\beta/(2\beta+1)})$. Further, assume that $\bar{b}_{ki_r} \neq \bar{b}_{kj_r}$ for $k \in \text{An}(i_r) \cap \text{An}(j_r)$ and $r = 1, \dots, h$. Then

$$\sqrt{k} \begin{pmatrix} \chi^{n,k}(i_1, j_1) - \chi(i_1, j_1) \\ \vdots \\ \chi^{n,k}(i_h, j_h) - \chi(i_h, j_h) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Pi),$$

where the asymptotic covariance matrix Π has for $r, q = 1, \dots, h$ the components (where we sum over all common ancestors)

$$\begin{aligned}\Pi((i_r, j_r), (i_q, j_q)) &= \sum \bar{b}_{ki_r} \wedge \bar{b}_{kj_r} \wedge \bar{b}_{ki_q} \wedge \bar{b}_{kj_q} \\ &\quad - \left(D(i_r) \sum \bar{b}_{ki_r} \wedge \bar{b}_{kj_r} \wedge \bar{b}_{ki_q} + D(j_r) \sum \bar{b}_{ki_r} \wedge \bar{b}_{kj_r} \wedge \bar{b}_{kj_q} \right. \\ &\quad \left. + D(i_q) \sum \bar{b}_{ki_q} \wedge \bar{b}_{kj_q} \wedge \bar{b}_{ki_r} + D(j_q) \sum \bar{b}_{ki_q} \wedge \bar{b}_{kj_q} \wedge \bar{b}_{kj_r} \right) \\ &\quad + \left(D(i_r)D(i_q)\chi(i_r, i_q) + D(j_r)D(i_q)\chi(j_r, i_q) \right. \\ &\quad \left. + D(i_r)D(j_q)\chi(i_r, j_q) + D(j_r)D(j_q)\chi(j_r, j_q) \right),\end{aligned}$$

with

$$D(i) := \sum_{k \in \text{An}(i) \cap \text{An}(j)} \bar{b}_{ki} \mathbf{1}_{\{\bar{b}_{ki} > \bar{b}_{kj}\}}.$$

Remark 6.2. In the recursive ML max-weighted model, for $i \in V$, $j \in \text{An}(i)$, by (2.7) we have $\bar{b}_{kj} < \bar{b}_{ki}$ for all $k \in \text{An}(i) \cap \text{An}(j)$. Hence, the last assumption in Theorem 6.1 is satisfied. \square

Proof of Theorem 6.1. The proof is inspired by the proof of Theorem 7.2.2 in de Haan and Ferreira [4]. For $r = 1, \dots, h$, we decompose $\sqrt{k}(\chi^{n,k}(i_r, j_r) - \chi(i_r, j_r))$ into

$$\sqrt{k}(\chi^{n,k}(i_r, j_r) - V^{n,k}(i_r, j_r)) + \sqrt{k}(V^{n,k}(i_r, j_r) - \chi(i_r, j_r)) =: I + II.$$

Asymptotic normality of II is shown in Proposition A.2 in a more general setting. The asymptotic covariance matrix Σ is given for the max-linear model by its coefficients

$$\Sigma((i_r, j_r), (i_q, j_q)) = \sum_{l=1}^d \bar{b}_{li_r} \wedge \bar{b}_{lj_r} \wedge \bar{b}_{li_q} \wedge \bar{b}_{lj_q},$$

and the summation reduces to $\text{An}(i_r) \cap \text{An}(j_r) \cap \text{An}(i_q) \cap \text{An}(j_q)$ for the recursive model. Since both I and II have an impact on the covariance matrix of the limiting normal distribution, we perform a Skorohod embedding to express the covariance between I and II in the limit. More precisely, we use the relation (6.2) and smoothness of the exponent measure μ_* from (2.5) at certain points to transfer the convergence result of II to I. This involves expressing the limit of I as a.s. limit to a Gaussian process.

We work with the function C defined for $\mathbf{x} > \mathbf{0} \in \mathbb{R}^d$ by

$$C(\mathbf{x}) := 2 - \mu_*([\mathbf{0}, \mathbf{x}^{-1}]^c) = 2 - \sum_{k=1}^d \bigvee_{l \in \text{De}(k)} x_l \bar{b}_{kl},$$

where $\mathbf{x}^{-1} := (x_1^{-1}, \dots, x_d^{-1})$. Note that $C(e_{i_r} \wedge e_{j_r}) = \chi(i_r, j_r)$ (cf. Remark 2.2), where the vector e_i is 1 at position i and ∞ elsewhere and the minimum is taken componentwise.

Set

$$U_{i,k:n} := 1 - F_i(X_{i,n-k:n}) \quad \text{and} \quad u_{i_r,j_r}^{n,k} := \frac{n}{k} U_{i_r,k:n} e_{i_r} \wedge \frac{n}{k} U_{j_r,k:n} e_{j_r}$$

and recall from (6.2) that $V^{n,k}(u_{i_r,j_r}^{n,k}) = \chi^{n,k}(i_r, j_r)$. By [17], Theorem 5, and a Skorohod embedding we obtain that

$$\sqrt{k}(C(u_{i_r,j_r}^{n,k}) - V^{n,k}(u_{i_r,j_r}^{n,k})) - W(u_{i_r,j_r}^{n,k}) \xrightarrow{n \rightarrow \infty} 0 \quad a.s., \quad (6.3)$$

where W is a Gaussian process with zero mean and known covariance structure given at our points of interest by

$$\text{Cov}(W(x_{i_r} e_{i_r} \wedge x_{j_r} e_{j_r}), W(y_{i_q} e_{i_q} \wedge y_{j_q} e_{j_q})) = \sum_{k=1}^d x_{i_r} \bar{b}_{ki_r} \wedge x_{j_r} \bar{b}_{kj_r} \wedge y_{i_q} \bar{b}_{ki_q} \wedge y_{j_q} \bar{b}_{kj_q}.$$

From Eq. (7.2.12) of [4] we obtain that

$$\sqrt{k}\left(\frac{n}{k} U_{i_r,k:n} - 1\right) - W(e_{i_r}) \xrightarrow{n \rightarrow \infty} 0 \quad a.s. \quad (6.4)$$

Since $\bar{b}_{ki_r} \neq \bar{b}_{kj_r}$, the partial derivatives of C w.r.t. x_{i_r} and x_{j_r} at the point $e_{i_r} \wedge e_{j_r}$ exist and are given by

$$\frac{\partial C(\mathbf{x})}{\partial x_{i_r}} \Big|_{\mathbf{x}=e_{i_r} \wedge e_{j_r}} = - \sum_{k \in \text{An}(i_r) \cap \text{An}(j_r)} \bar{b}_{ki_r} \mathbf{1}_{\{\bar{b}_{ki_r} > \bar{b}_{kj_r}\}} =: -D(i_r),$$

we find by Cramér's delta method and (6.4) that

$$\sqrt{k}\left(C(u_{i_r, j_r}^{n,k}) - C(e_{i_r} \wedge e_{j_r})\right) + D(i_r)W(e_{i_r}) + D(j_r)W(e_{j_r}) \xrightarrow{n \rightarrow \infty} 0 \quad a.s.$$

By continuity of the paths of a Gaussian process,

$$\left|W(u_{i_r, j_r}^{n,k}) - W(e_{i_r} \wedge e_{j_r})\right| \xrightarrow{n \rightarrow \infty} 0 \quad a.s.,$$

and we conclude from (6.3) that

$$\sqrt{k}\left(\chi(i_r, j_r) - \chi^{n,k}(i_r, j_r)\right) - W(e_{i_r} \wedge e_{j_r}) + D(i_r)W(e_{i_r}) + D(j_r)W(e_{j_r}) \xrightarrow{n \rightarrow \infty} 0 \quad a.s.$$

Slutzky's Theorem applies, and we obtain asymptotic normality of $I + II$ with the covariance matrix Π as stated in the theorem. \square

Acknowledgements

We thank Steffen Lauritzen and Jonas Peters for interesting discussions and invaluable help with the graph terminology. MO and NG thank the International Graduate School of Science and Engineering (IGSSE) of the Technical University of Munich for support.

References

- [1] A. V. Aho, M. R. Garey, and J. D. Ullman. The transitive reduction of a directed graph. *SIAM Journal on Computing*, 1(2):131–137, 1972.
- [2] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes: Theory and Applications*. Wiley, Chichester, 2004.
- [3] K. Bollen. *Structural Equations with Latent Variables*. Wiley, New York, 1989.
- [4] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, New York, 2006.
- [5] R. Diestel. *Graph Theory*. Graduate texts in mathematics; 173. Springer, 4 edition, 2010.
- [6] R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, Cambridge, 2010.
- [7] D. Edwards. *Introduction to Graphical Modelling*. Springer, New York.
- [8] N. Gissibl and C. Klüppelberg. Max-linear models on directed acyclic graphs. Available at arXiv:1512.07522, 2015.
- [9] D. Heckerman and D. Geiger. Likelihoods and parameter priors for Bayesian networks. Technical Report MSR-TR-95-54, Microsoft Research, 1995.
- [10] D. Koller and N. Friedman. *Probabilistic Graphical Models*. MIT Press, 2009.
- [11] S. L. Lauritzen. Causal inference from graphical models. In *Complex Stochastic Systems*. Chapman and Hall/CRC, Boca Raton.
- [12] S. L. Lauritzen. *Graphical Models*. Oxford University Press, Oxford, 1996.
- [13] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2nd edition, 2009.

- [14] J. Peters. *Restricted Structural Equation Models for Causal Inference*. Dissertation, ETH Zurich, 2012.
- [15] S. I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer, New York, 2007.
- [16] S. I. Resnick. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York, 2007.
- [17] R. Schmidt and U. Stadtmüller. Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics*, 33(2):307–335, 2006.

A Consistency and asymptotic normality of the empirical estimator

We assume the distributional situation as in Section 2 and Section 6; i.e., the data $\mathbf{X} = (X_1, \dots, X_d)$ is (marginally) regularly varying with index $\alpha > 0$ and has (after standardization of the marginals) exponent measure $\mu_*([\mathbf{0}, \mathbf{x}]^c) = \sum_{k=1}^d \bigvee_{i \in \text{De}(k)} \frac{b_{ki}^\alpha}{\sum_{l=1}^d b_{li}^\alpha} x_i^{-1}$ (cf. Remark 2.2). The following is a consequence of (3.2).

Lemma A.1. *Assume $1 \leq r, q \leq h$. Then there exists some positive definite matrix Σ with components for $1 \leq r, q \leq h$ given by*

$$\Sigma((i_r, j_r), (i_q, j_q)) := \lim_{x \rightarrow 1} \frac{1}{1-x} \left(\mathbb{P}(F_{i_r}(X_{i_r}) > x, F_{j_r}(X_{j_r}) > x, F_{i_q}(X_{i_q}) > x, F_{j_q}(X_{j_q}) > x) - \mathbb{P}(F_{i_r}(X_{i_r}) > x, F_{j_r}(X_{j_r}) > x) \mathbb{P}(F_{i_q}(X_{i_q}) > x, F_{j_q}(X_{j_q}) > x) \right).$$

The following Lemma proves asymptotic normality of the empirical matrix estimator for χ under general conditions. Theorem 6.1 is a special case specifying Σ for the estimator $V^{n,k}$ in a max-linear model.

Proposition A.2. *Let $(i_1, j_1), \dots, (i_h, j_h)$ be arbitrary node pairs in \mathcal{D} . Suppose that for $k = 1, \dots, h$ and $\beta > 0$,*

$$x \mathbb{P}\left(F_{i_k}(X_{i_k}) > 1 - \frac{1}{x}, F_{j_k}(X_{j_k}) > 1 - \frac{1}{x}\right) = \chi(i_k, j_k) + O(x^{-\beta}). \quad (\text{A.1})$$

Let $k = k(n) \rightarrow \infty$ as $n \rightarrow \infty$ and $k = o(n^{2\beta/(2\beta+1)})$. Then we have

$$\sqrt{k} \begin{pmatrix} V_{i_1, j_1}^{n,k} - \chi(i_1, j_1) \\ \vdots \\ V_{i_h, j_h}^{n,k} - \chi(i_h, j_h) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma) \quad (\text{A.2})$$

with Σ specified in Lemma A.1

Proof. Our proof is based on ideas of the proof of Theorem 5 of Schmidt and Stadtmüller [17]. Recall that $V_{i_1, j_1}^{n,k} = V^{n,k}(e_i + e_j)$ in (6.1). The first goal is to show that

$$\sqrt{k} \begin{pmatrix} V_{i_1, j_1}^{n,k} - \frac{n-k}{k} \mathbb{P}(F_{i_1}(X_{i_1}) > \frac{n-k}{n}, F_{j_1}(X_{j_1}) > \frac{n-k}{n}) \\ \vdots \\ V_{i_h, j_h}^{n,k} - \frac{n-k}{k} \mathbb{P}(F_{i_h}(X_{i_h}) > \frac{n-k}{n}, F_{j_h}(X_{j_h}) > \frac{n-k}{n}) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma). \quad (\text{A.3})$$

By the Cramér-Wold device, it suffices to show that for every vector $t = (t_1, \dots, t_h)^T \in \mathbb{R}^h$,

$$\sqrt{k} \sum_{r=1}^h t_r \left(V_{i_r, j_r}^{n, k} - \frac{n}{k} \mathbb{P} \left(F_{i_r}(X_{i_r}) > \frac{n-k}{n}, F_{j_r}(X_{j_r}) > \frac{n-k}{n} \right) \right) \xrightarrow{d} \mathcal{N}(0, t^T \Sigma t). \quad (\text{A.4})$$

$$\sqrt{k} \sum_{r=1}^h \frac{t_r}{k} \left(\mathbf{1}_{\{F_{i_r}(X_{m, i_r}), F_{j_r}(X_{m, j_r}) > \frac{n-k}{n}\}} - \mathbb{P} \left(F_{i_r}(X_{i_r}) > \frac{n-k}{n}, F_{j_r}(X_{j_r}) > \frac{n-k}{n} \right) \right) \xrightarrow{d} \mathcal{N}(0, t^T \Sigma t).$$

In order to show this, we rewrite the lhs of (A.4) as

$$\begin{aligned} & \sqrt{k} \sum_{r=1}^h t_r \left(V_{i_r, j_r}^{n, k} - \frac{n}{k} \mathbb{P} \left(F_{i_r}(X_{i_r}) > \frac{n-k}{n}, F_{j_r}(X_{j_r}) > \frac{n-k}{n} \right) \right) \\ &= \sqrt{k} \sum_{r=1}^h t_r \frac{1}{k} \sum_{m=1}^n \left(\mathbf{1}_{\{F_{i_r}(X_{i_r, m}) > \frac{n-k}{n}, F_{j_r}(X_{j_r, m}) > \frac{n-k}{n}\}} - \mathbb{P} \left(F_{i_r}(X_{i_r}) > \frac{n-k}{n}, F_{j_r}(X_{j_r}) > \frac{n-k}{n} \right) \right) \\ &= \sum_{m=1}^n \sum_{r=1}^h t_r \frac{1}{\sqrt{k}} \left(\mathbf{1}_{\{F_{i_r}(X_{i_r, m}) > \frac{n-k}{n}, F_{j_r}(X_{j_r, m}) > \frac{n-k}{n}\}} - \mathbb{P} \left(F_{i_r}(X_{i_r}) > \frac{n-k}{n}, F_{j_r}(X_{j_r}) > \frac{n-k}{n} \right) \right) \\ &=: \sum_{m=1}^n X_{n, m}, \end{aligned}$$

where the random variables

$$X_{n, m} := \sum_{r=1}^h t_r \frac{1}{\sqrt{k}} \left(\mathbf{1}_{\{F_{i_r}(X_{i_r, m}) > \frac{n-k}{n}, F_{j_r}(X_{j_r, m}) > \frac{n-k}{n}\}} - \mathbb{P} \left(F_{i_r}(X_{i_r}) > \frac{n-k}{n}, F_{j_r}(X_{j_r}) > \frac{n-k}{n} \right) \right),$$

for $m = 1, \dots, n$ are independent, but depend on n . Hence, we are in a situation to apply a Lindeberg-Feller type Central Limit Theorem for triangular arrays (Durrett [6], Theorem 3.4.5.) to establish $\sum_{m=1}^n X_{n, m} \xrightarrow{d} \mathcal{N}(0, t^T \Sigma t)$. We have

$$\begin{aligned} E(X_{n, m}) &= 0, \\ E(X_{n, m}^2) &= \sum_{r=1}^h \sum_{q=1}^h t_r t_q \frac{1}{k} \text{Cov} \left(\mathbf{1}_{\{F_{i_r}(X_{i_r, m}) > \frac{n-k}{n}, F_{j_r}(X_{j_r, m}) > \frac{n-k}{n}\}}, \mathbf{1}_{\{F_{i_q}(X_{i_q, m}), F_{j_q}(X_{j_q, m}) > \frac{n-k}{n}\}} \right). \end{aligned}$$

Hence, with the transformation $\frac{n-k}{n} \mapsto x$ in the third equality, we obtain by (3.2),

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{m=1}^n E(X_{n, m}^2) &= \lim_{n \rightarrow \infty} n E(X_{n, 1}^2) \\ &= \lim_{x \rightarrow 1} \sum_{r=1}^h \sum_{q=1}^h t_r t_q \frac{1}{1-x} \left(\mathbb{P} \left(F_{i_r}(X_{i_r}) > x, F_{j_r}(X_{j_r}) > x, F_{i_q}(X_{i_q}) > x, F_{j_q}(X_{j_q}) > x \right) \right. \\ &\quad \left. - \mathbb{P} \left(F_{i_r}(X_{i_r}) > x, F_{j_r}(X_{j_r}) > x \right) \mathbb{P} \left(F_{i_q}(X_{i_q}) > x, F_{j_q}(X_{j_q}) > x \right) \right) \\ &= t^T \Sigma t \end{aligned}$$

by Lemma A.1. Furthermore, we need to prove that for all $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \sum_{m=1}^n E(X_{n, m}^2 \mathbf{1}_{\{|X_{n, m}| > \varepsilon\}}) = 0. \quad (\text{A.5})$$

From the definition of $X_{n,m}$ we find that $X_{n,m}^2$ is of order $\frac{1}{k}$. For arbitrary $\varepsilon > 0$, by Markov's inequality,

$$\begin{aligned} E(\mathbf{1}_{\{|X_{n,1}|>\varepsilon\}}) &\leq \varepsilon^{-1} E(|X_{n,1}|) \\ &\leq \varepsilon^{-1} \sum_{r=1}^h t_r \frac{2}{\sqrt{k}} \left(\mathbb{P}\left(F_{i_r}(X_{i_r}) > \frac{n-k}{n}, F_{j_r}(X_{j_r}) > \frac{n-k}{n}\right) - \mathbb{P}^2\left(F_{i_r}(X_{i_r}) > \frac{n-k}{n}, F_{j_r}(X_{j_r}) > \frac{n-k}{n}\right) \right). \end{aligned}$$

By (3.2), $\mathbb{P}(F_{i_r}(X_{i_r}) > \frac{n-k}{n}, F_{j_r}(X_{j_r}) > \frac{n-k}{n}) = O(\frac{k}{n})$ as $n \rightarrow \infty$. Thus, $\mathbb{P}(|X_{n,1}| > \varepsilon)$ is smaller order than $\frac{\sqrt{k}}{n}$ as $n \rightarrow \infty$. Hence,

$$\sum_{m=1}^n E(X_{n,m}^2 \mathbf{1}_{\{|X_{n,m}|>\varepsilon\}}) = nE(X_{n,1}^2 \mathbf{1}_{\{|X_{n,1}|>\varepsilon\}}) \leq O\left(\frac{n}{k}\right) O\left(\frac{\sqrt{k}}{n}\right) \xrightarrow{n \rightarrow \infty} 0,$$

which shows (A.5). Hence, the assumptions of the Lindeberg-Feller Central Limit Theorem are satisfied and (A.3) holds. If we can show that

$$\sqrt{k} \left(\chi(i_k, j_k) - \frac{n}{k} \mathbb{P}\left(F_{i_k}(X_{i_k}) > \frac{n-k}{n}, F_{j_k}(X_{j_k}) > \frac{n-k}{n}\right) \right) \xrightarrow{n \rightarrow \infty} 0,$$

then (A.2) follows from (A.3) and Slutsky's Theorem.

Indeed, invoking (A.1) and the rate of k , we obtain

$$\begin{aligned} &\sqrt{k} \left(\chi(i_k, j_k) - \frac{n}{k} \mathbb{P}\left(F_{i_k}(X_{i_k}) > \frac{n-k}{n}, F_{j_k}(X_{j_k}) > \frac{n-k}{n}\right) \right) \\ &= \sqrt{k} O((k/n)^\beta) = o(n^{\beta/(2\beta+1)}) o(n^{2\beta^2/(2\beta+1)} n^{-1}) \xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

and finish the proof. \square